

学修番号 17890509

修士論文

マルチタスク学習を用いた日本語述語項構造解析

大森 光

2019年2月22日

首都大学東京大学院
システムデザイン研究科 情報通信システム学域

大森 光

審査委員：

小町 守 准教授 (主指導教員)

山口 亨 教授 (副指導教員)

高間 康史 教授 (副指導教員)

マルチタスク学習を用いた日本語述語項構造解析*

大森 光

修論要旨

述語項構造解析とはテキスト中に存在する述語とその項との意味構造を解析するタスクである。「誰が」「何を」「誰に」「どうした」のように文の構造を整理することは、機械翻訳や含意関係認識など複雑な文章の解析を必要とする処理のために有用である。述語項構造解析は動詞や形容詞といった述語を対象にし、その項を解析するタスクであるが、文中には名詞でも項を持つものが多く存在する。例えば、「報告」のようなサ変名詞や「救い」のような動詞から派生した名詞がこれにあたり、このような名詞を事態性名詞と呼ぶ。NAIST テキストコーパスでは、述語と事態性名詞はどちらも必須格（ガ格、ヲ格、ニ格）の項を持つ。また飯田ら (2006) によると、述語は同一文節内に項を持つことはほぼないが、事態性名詞の場合はヲ格とニ格の項の約半分が同一文節内に出現する。このことから、事態性名詞の項構造解析と述語項構造解析は関係性が高いが別のタスクであるといえる。先行研究では、機械学習を用いた述語項構造解析の研究が盛んに行われてきた。しかしこれらのほとんどが述語を対象に項の解析を行っており、事態性名詞を対象とした研究は少ない。文章の意味構造を整理し、正しく文脈解析を行ううえで、述語のみを対象にした項構造解析は不十分である。

そこで本研究では、述語項構造解析と事態性名詞の項構造解析をマルチタスク学習するモデルを提案する。マルチタスク学習は含意関係認識や意味役割付与など自然言語処理の様々な分野で適用されており、精度の向上が報告されている学習手法である。マルチタスク学習の特徴として、複数のタスクを同時に解くことで学習データが増え、各タスクに含まれるノイズに対して頑健に学習できることが挙げられる。また、タスク間で共通する知識を獲得することで、モデルをより汎化できるのも利点の一つである。NAIST テキストコーパスは機械翻訳などで用いられてい

*首都大学東京大学院 システムデザイン研究科 情報通信システム学域 修士論文, 学修番号 17890509, 2019年2月22日.

る大規模なデータセットと比較すると小規模であり，NAIST テキストコーパスにおける事態性名詞のデータ数は述語のデータ数の約 3 分の 1 であることから，特に事態性名詞の項構造解析においてマルチタスク学習が効果的であることが期待できる．この提案モデルでは end-to-end な多層双方向 Recurrent Neural Network (RNN) がベースであり，入力層と RNN 層と出力層においてタスク間で共通の知識とタスク特有の知識を区別するアーキテクチャーを有する．入力層では，タスク独自の単語ベクトルを学習することで，表層形が同じであっても文脈の異なる述語と事態性名詞の違いを区別する．RNN 層では，タスク共有の RNN の上にタスク特有の RNN を階層的に重ねるニューラルネットワークを構築することで，共有の RNN でタスク共有の知識表現を学習し，タスク特有の RNN でそれぞれのタスクに調整する．出力層では，タスク共有の層とタスク特有の層に分けることで，述語と事態性名詞の項の出現位置の違いを考慮するように学習する．

本論文では日本語述語項構造解析において一般的なベンチマークとして使われる NAIST テキストコーパスを用いて実験を行った．実験の結果，提案モデルがマルチタスク学習を用いないベースラインモデルと比較して，全体の F 値で 0.29 ポイントの精度向上を示した．また，文内述語項構造解析タスクにおいて最高精度を達成している先行研究のモデルを 0.67 ポイント上回る解析精度を示した．

本論文の貢献は以下の 3 つである．

1. 初めてマルチタスク学習を用いて述語項構造解析と事態性名詞の項構造解析を行い，双方のタスクで解析精度が向上することを示した．
2. 統語情報を素性として加えることによって，複数の述語項関係を考慮しない単純なモデルが文内述語項構造解析において最高精度を達成した．
3. 初めて深層学習を用いて事態性名詞の項構造解析を行った．

本稿の構成は以下のようになっている．第 1 章では本研究の概要について述べる．第 2 章では日本語述語項構造解析と意味役割付与についての関連研究について述べる．第 3 章ではタスク設定と end-to-end モデルの概要及び素性について述べる．第 4 章では提案手法である述語項構造解析と事態性名詞の項構造解析のマルチタスク学習モデルについて説明する．第 5 章ではデータセットと実験設定，実験結果について述べる．第 6 章ではベースラインと提案モデルの出力結果を比較し，各

モデルの効果を分析する。最後に，第 7 章で本研究のまとめと今後の展望について述べる。

Multi-task Learning for Japanese Predicate Argument Structure Analysis*

Hikaru Omori

Abstract

An *event-noun* is a noun that has an argument structure similar to a predicate. Recent works, including those considered state-of-the-art, ignore event-nouns or build a single model for solving both Japanese predicate argument structure analysis (PASA) and event-noun argument structure analysis (ENASA). However, because there are interactions between predicates and event-nouns, it is not sufficient to target only predicates. To address this problem, we present a multi-task learning method for PASA and ENASA. Our multi-task models improved the performance of both tasks compared to a single-task model by sharing knowledge from each task. Moreover, in PASA, our models achieved state-of-the-art results in overall F1 scores on the NAIST Text Corpus. In addition, this is the first work to employ neural networks in ENASA.

*Master's Thesis, Department of Information and Communication Systems, Graduate School of System Design, Tokyo Metropolitan University, Student ID 17890509, February 22, 2019.

目次

図目次	vii
第 1 章 はじめに	1
第 2 章 関連研究	3
2.1 日本語述語項構造解析と事態性名詞の項構造解析のアプローチ	3
2.2 ニューラルネットワークを用いた述語項構造解析	3
2.3 意味役割付与	4
第 3 章 述語と事態性名詞の項構造解析	6
3.1 タスク設定	6
3.2 シングルタスクモデル	6
3.2.1 入力層	7
述語（事態性名詞）と項	7
位置	7
係り受け	7
事態性	8
タスクラベル	8
3.2.2 RNN 層	8
3.2.3 出力層	9
第 4 章 マルチタスクモデル	10
4.1 マルチ入力層	11
4.2 マルチ RNN 層	11

4.3	マルチ出力層	12
第 5 章	実験	13
5.1	データセットと実験設定	13
5.2	ハイパーパラメータ	13
5.3	実験結果	14
5.3.1	述語項構造解析	14
5.3.2	事態性名詞の項構造解析	16
第 6 章	分析	17
第 7 章	おわりに	19
	発表リスト	20
	謝辞	21
	参考文献	22

目次

1.1	述語項構造解析と事態性名詞の項構造解析の例。矢印は係り受け 関係を示す。	1
3.1	end-to-end ベースラインモデル	7
4.1	提案モデル: (a) Single, (b) Multi-input, (c) Multi-RNN, (d) Multi-output, (e) Multi-ALL.	10

第 1 章 はじめに

述語項構造解析とはテキスト中に存在する述語とその項との意味構造を解析するタスクである。「誰が」「何を」「誰に」「どうした」のように文の構造を整理することは、機械翻訳や含意関係認識など複雑な文章の解析を必要とする処理のために有用である。述語項構造解析は動詞や形容詞といった述語を対象にし、その項を解析するタスクであるが、文中には名詞でも項を持つものが多く存在する。例えば、「報告」のようなサ変名詞や「救い」のような動詞から派生した名詞がこれにあたり、このような名詞を事態性名詞と呼ぶ。

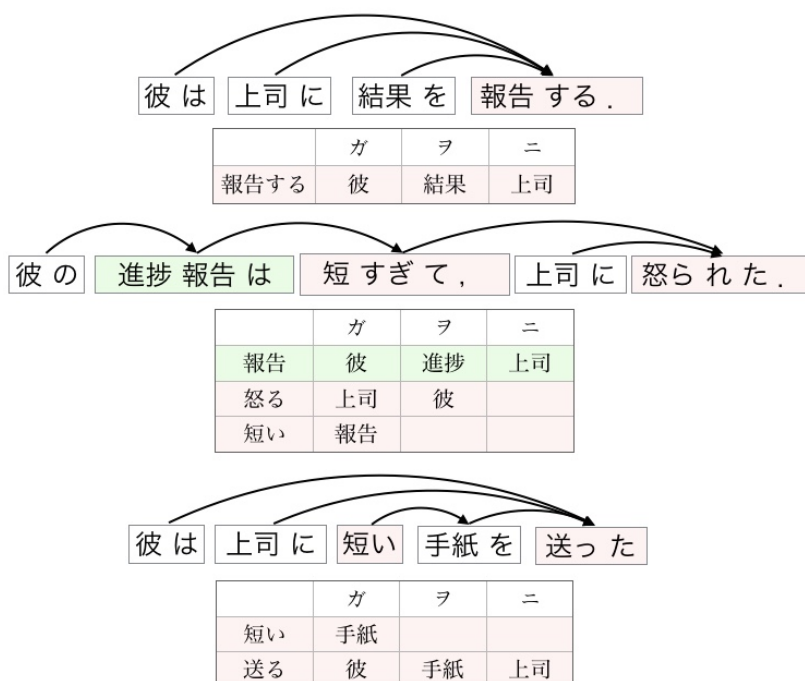


図 1.1: 述語項構造解析と事態性名詞の項構造解析の例。矢印は係り受け関係を示す。

述語項構造解析と事態性名詞の項構造解析の例を図 1.1 に示す。NAIST テキストコーパス [1] では、述語と事態性名詞はどちらも必須格（ガ格，ヲ格，ニ格）の項を持つ。また Iida ら [1] によると、述語は同一文節内に項を持つことはほぼないが、事態性名詞の場合はヲ格とニ格の項の約半分が同一文節内に出現する。このこ

とから、事態性名詞の項構造解析と述語項構造解析は関係性が高いが別のタスクであるといえる。

先行研究では、機械学習を用いた述語項構造解析の研究が盛んに行われてきた。しかしこれらのほとんどが述語を対象に項の解析を行っており、事態性名詞を対象とした研究 [2, 3] は少ない。文章の意味構造を整理し、正しく文脈解析を行ううえで、述語のみを対象にした項構造解析は不十分である。そこで我々は、述語項構造解析と事態性名詞の項構造解析をマルチタスク学習するモデルを提案する。このモデルは先行研究で用いられた end-to-end な多層双方向 Recurrent Neural Network (RNN) がベースであり、入力層と RNN 層と出力層においてタスク間で共通の知識とタスク特有の知識を区別するネットワークを有する。さらに、この提案モデルは先行研究と同様に述語と項の統語関係を考慮する。我々は日本語述語項構造解析において一般的なベンチマークとして使われる NAIST テキストコーパスを用いて実験を行った。

本稿の貢献は以下の 3 つである。

1. 初めてマルチタスク学習を用いて述語項構造解析と事態性名詞の項構造解析を行い、双方のタスクで解析精度が向上することを示した。
2. 統語情報を素性として加えることによって、複数の述語項関係を考慮しない単純なモデルが文内述語項構造解析において最高精度を達成した。
3. 初めて深層学習を用いて事態性名詞の項構造解析を行った。

本稿ではまず、第 2 章で日本語述語項構造解析と意味役割付与についての関連研究について述べる。第 3 章ではタスク設定と end-to-end モデルの概要及び素性について述べる。次に、第 4 章では提案手法である述語項構造解析と事態性名詞の項構造解析のマルチタスク学習モデルについて説明する。第 5 章ではデータセットと実験設定、実験結果について述べる。第 6 章ではベースラインと提案モデルの出力結果を比較し、各モデルの効果を分析する。最後に、第 7 章で本研究のまとめと今後の展望について述べる。

第2章 関連研究

2.1 日本語述語項構造解析と事態性名詞の項構造解析のアプローチ

日本語述語項構造解析では機械学習に基づく手法が多く研究されてきた。伝統的な手法は述語の必須格（ガ，ヲ，ニ）毎に異なるモデルを構築する pointwise なアプローチをとる。Taira ら [3] は決定リストと Support Vector Machines を用いて格毎に異なる素性の重みを学習するモデルを提案した。Imamura ら [4] は最大エントロピー法を大規模な新聞データから訓練された言語モデルに適用する手法を提案した。Hayashibe ら [5] は項と述語の位置関係によって異なるモデルを作成し、トーナメントモデルによって最尤のモデル出力を決定する手法を提案した。それに対し、文毎に最適化を行い、複数の述語と項を同時に解析する joint なアプローチが pointwise なアプローチより高い解析精度を示している。Yoshikawa ら [6] は Markov Logic Network を用いて複数の述語項関係の間にある依存関係を考慮するモデルを提案した。Ouchi ら [7] はグラフを用いて文内にある全ての述語項関係を同時に考慮し、最適化を行った。

これらの研究は Taira らを除いて述語の解析に焦点を当てたものであるが、一方で事態性名詞について焦点を当てた研究も存在する。Komachi ら [2] は事態性名詞の項構造解析を事態性判別タスクと項同定タスクに分解し、語彙統語パターンを用いた手法を提案した。これまでの事態性名詞の先行モデルは pointwise アプローチであり、Taira ら [3] は述語と事態性名詞の解析を、Komachi ら [2] は事態性名詞のみの解析を単一のモデルで行っているため、それぞれのタスクで有効な素性が異なることを考慮していなかった。我々のモデルでは3つの必須格を同時に推定して最適化を行っており、またマルチタスク学習によってタスク間で共通の特徴とタスク特有の特徴を区別して学習することができる。

2.2 ニューラルネットワークを用いた述語項構造解析

述語項構造解析でもニューラルネットワークを用いた手法が従来の手法よりも高い解析精度を示している。Shibata ら [8] は Ouchi ら [7] のモデルの一部に Feed

Forward Neural Networks を適用した。Matsubayashi と Inui [9] は統語関係埋め込みベクトルなどを使って述語とその項から構成される局所的な素性を表現し、複数の述語とその項の関係を考慮しない局所的なモデルでも高い精度を達成できることを示した。

さらに end-to-end なモデルもいくつか提案されている。Ouchi ら [10] は Zhou と Xu [11] が提案したモデルをベースに end-to-end モデルを構築し、Grid RNN を用いて複数の述語を同時に考慮できる手法を提案した。通常の RNN が時間方向の接続のみに対し、Grid RNN はセルを多次元の格子状に配置することができるため、データ系列方向にも接続が可能である。複数のデータ系列に相互関係がある場合は有効な手法であるといえる。Matsubayashi と Inui [12] は複数の述語の相互関係を直接捉える目的で Ouchi ら [10] のモデルに self-attention を組み合わせる手法を提案した。self-attention とはデータ系列のどの部分を重要視するかを考慮する attention 機構の 1 種である。RNN が直前の隠れ層のみを参照するのに対し、self-attention では全ての隠れ層を考慮することができ、複数の self-attention を多層かつ並列に結合できるため、長距離の依存関係を捉えることに優れている。この提案モデルでは、特に述語と係り受け関係にない項の解析において解析精度が向上し、文内述語項構造解析タスクにおいて最高精度を達成することを示した。

2.3 意味役割付与

述語項構造解析と似たタスクに意味役割付与があり、このタスクでは A0（動作主）や TMP（時間）等といった深層レベルのラベルを付与することで述語項間の意味関係をより詳しく整理する。近年、英語の意味役割付与ではニューラルネットワークを用いた end-to-end なモデルが高い解析精度を示している。Zhou と Xu [11] は 8 層の双方向 Long Short-Term Memory (LSTM) に Conditional Random Field を組み合わせたモデルを提案した。He ら [13] は Highway LSTM を用いたモデルを構築し制約付きの decoding を行った。Tan ら [14] は self-attention を用いることで、長距離の依存関係を直接的に捉えることができることを示した。Strubell ら [15] は self-attention を用いたモデルをベースに、係り受け解析、品詞タグ付け、述語検出、意味役割付与の 4 つのタスクでマルチタスク学習を行った。

Ouchi ら [16] は双方向 LSTM を用いたスパンベースのモデルで、全てのスパンを各ラベルごとにスコア付けし、制約付きの decoding を行うことで最高精度を達成することを示した。事態性名詞に焦点を当てた研究としては、Gerber と Chai [17] が頻度の高い 10 種類の事態性名詞に対して、Pointwise Mutual Information などを素性として利用し、ロジスティック回帰モデルによって意味役割を識別するモデルを提案した。

中国語の意味役割付与においても、LSTM を用いた手法が高い解析精度を示している [18, 19, 20, 21, 22]。事態性名詞の研究に関しては、Li ら [23] が動詞意味役割付与において有効な素性を名詞意味役割付与に組み合わせることで精度が向上することを示した。彼らは動詞意味役割付与の精度向上は示していないが、我々はマルチタスク学習を利用することで述語項構造解析と事態性名詞の項構造解析の両方のタスクに対して解析精度が向上することを示した。

第 3 章 述語と事態性名詞の項構造解析

3.1 タスク設定

日本語述語（事態性名詞）項構造解析とは、ある述語（事態性名詞）に対する項を抽出しガ格，ヲ格，ニ格の 3 つの格ラベルを付与するタスク [1] である。項は述語（事態性名詞）の位置によって 4 つのカテゴリ [3] に大別される。

Dep 項が述語（事態性名詞）に係る，もしくは述語（事態性名詞）が項に係る。

Zero ゼロ照応によって項が省略され，項と述語（事態性名詞）が同一文内にあるが，直接係り受け関係にない。

Inter-Zero ゼロ照応によって項が省略され，項と述語（事態性名詞）が同一文内にあるが，直接係り受け関係にない。

Bunsetsu 項と述語（事態性名詞）が同一文節内にある。

単語単位に分かち書きされた文 $w = w_1, w_2, \dots, w_T$ と述語（事態性名詞）の位置 $p = p_1, p_2, \dots, p_q$ が入力として与えられる。Iida ら [24] や Imamura ら [4] や Sasano ら [25] は Inter-Zero の解析を行っているが，これは他のカテゴリと違い文書全体を探索しなければならないため，解析がより困難である。本研究では既存研究 [7, 10, 9, 12] に従い，項と述語（事態性名詞）が同一文内にある事例のみを解析対象とする。また Ouchi ら [10] や Matsubayashi と Inui [12] にならい，Bunsetsu カテゴリに該当する項は述語項構造解析の評価対象から除外する。

3.2 シングルタスクモデル

我々は，先行研究 [11, 10, 12] で用いられた end-to-end な多層双方向 RNN をベースにしたマルチタスク項構造解析モデルを提案する。Single-Task モデルを図 3.1 に示す。

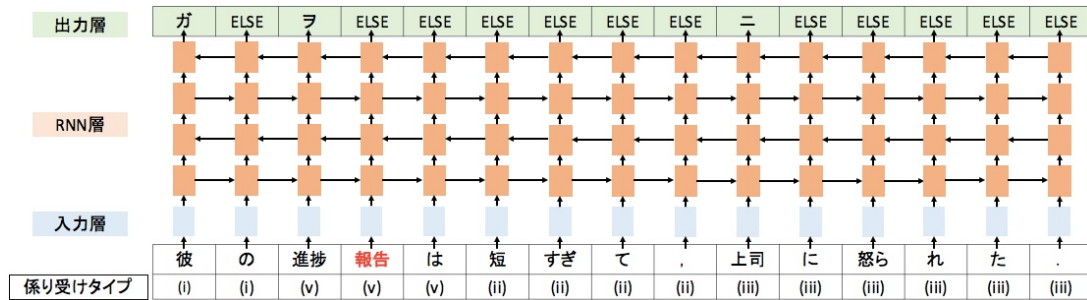


図 3.1: end-to-end ベースラインモデル

3.2.1 入力層

単語 $w_t \in [w_1, \dots, w_T]$ は入力層で素性ベクトル $\mathbf{x}_t \in [\mathbf{x}_1, \dots, \mathbf{x}_T]$ に変換される。素性ベクトル \mathbf{x}_t は以下のように定義される。

$$\mathbf{x}_t = \mathbf{x}_t^{\text{as}} \oplus \mathbf{x}_t^{\text{posi}} \oplus \mathbf{x}_t^{\text{dep}} \oplus \mathbf{x}_t^{\text{type}} \oplus \mathbf{x}_p^{\text{task}} \quad (3.21)$$

\oplus はベクトルの結合を表す。

■述語（事態性名詞）と項 述語（事態性名詞） w_p と項候補の単語 w_t は単語埋め込み行列によってベクトル $\mathbf{x}_t^{\text{as}} \in \mathbb{R}^{2d_w}$ にそれぞれ変換される。 d_w は単語分散表現の次元数である。

■位置 述語（事態性名詞）と単語の位置関係を示す素性であり、述語（事態性名詞）のインデックスから単語のインデックスを引いた値をとる。単語単位 p_t^{word} と文節単位 p_t^{bunsetsu} の 2 種類があり、単語位置埋め込み行列、文節位置埋め込み行列によって単語位置ベクトル $\mathbf{p}_t^{\text{word}} \in \mathbb{R}^{d_p}$ 、文節位置ベクトル $\mathbf{p}_t^{\text{bunsetsu}} \in \mathbb{R}^{d_p}$ にそれぞれ変換される。この 2 つのベクトルを結合し、位置ベクトル $\mathbf{x}_t^{\text{posi}} \in \mathbb{R}^{2d_p}$ を得る。

■係り受け 述語（事態性名詞）と項候補の係り受け関係を示す素性 d_t である。係り受け関係のタイプは 5 種類に設定した：

- i). 項が述語（事態性名詞）に係る。
- ii). 述語（事態性名詞）が項に係る。
- iii). 項と述語（事態性名詞）が係り受け関係にない。

- iv). 述語が項と同一文節内にある.
- v). 事態性名詞が項と同一文節内にある.

d_t は係り受け関係埋め込み行列によって係り受けベクトル $\mathbf{x}_t^{\text{dep}} \in \mathbb{R}^{d_a}$ に変換される. 図 1.1 の文を例に素性の作り方を図 3.1 に示す.

先行研究でも係り受け関係を用いた素性が使用されている. Imamura ら [4] は述語と項の間に係り受け関係があるか否かの 2 値素性を用いた. 我々は事態性名詞に適応させるために, より細かく係り受け関係を分類した. Matsubayashi と Inui [9] は統語関係埋め込みベクトルを用いて述語項間の関係を表現した. 対照的に, 我々は述語と事態性名詞を区別するためにそれぞれで異なる素性を定義し, これらの潜在的な構造を見つけるように学習を行った.

■事態性 この素性は文中の全ての述語 (事態性名詞) の位置にフラグを立てる 2 値素性ベクトルである. この素性は Matsubayashi と Inui [12] に影響されたものである. この素性の目的は述語が項になるのを防ぎ, 一部の事態性名詞が項になるのを助けることである. 項候補タイプベクトル $\mathbf{x}_t^{\text{type}} \in \mathbb{R}^2$ は項候補が述語であれば $[0,1]$, 項候補が事態性名詞であれば $[1,0]$, そうでなければ $[0,0]$ を示す.

■タスクラベル タスクが述語項構造解析であれば 1, 事態性の項構造解析であれば 0 を示す 2 値素性ベクトル $\mathbf{x}_p^{\text{task}} \in \mathbb{R}^1$ である.

3.2.2 RNN 層

RNN に Gated Recurrent Unit (GRU) [26] を用い, RNN 層は L 層の Stacked Bi-directional Gated Recurrent Unit (Stacked Bi-GRU) で構成する. さらに先行研究 [10, 12] に従って, residual connections [27] を行った. 各タイムステップ t において, $l \in [1, \dots, L]$ 層目の隠れ層ベクトル $\mathbf{h}_t^l \in \mathbb{R}^{d_h}$ は以下のように計算される.

$$\mathbf{h}_t^l = \begin{cases} g^l(\mathbf{h}_t^{l-1}, \mathbf{h}_{t-1}^l) & (l = \text{odd}) \\ g^l(\mathbf{h}_t^{l-1}, \mathbf{h}_{t+1}^l) & (l = \text{even}) \end{cases} \quad (3.22)$$

$g^l(\cdot)$ は l 層目の GRU の処理を表す. また $\mathbf{h}_t^0 = \mathbf{x}_t$ である.

3.2.3 出力層

その後出力層に各隠れ層のベクトル \mathbf{h}_t^L を入力し，得られたベクトルをソフトマックス関数を用いて正規化したものを出力ベクトル \mathbf{o}_t とする．

$$\mathbf{o}_t = \text{softmax}(\mathbf{W}_o \mathbf{h}_t^L + \mathbf{b}_o) \quad (3.23)$$

$\mathbf{W}_o \in \mathbb{R}^{4 \times d_h}$ はパラメータ行列， $\mathbf{b}_o \in \mathbb{R}^4$ はバイアス項である．出力ベクトル \mathbf{o}_t の各要素は，項候補が述語のガ格，ヲ格，ニ格であるか，必須格でないか (ELSE) を表す確率値となっており，確率値最大のラベルを予測ラベル y_t として出力する．モデルパラメータはクロスエントロピー誤差関数に基づいて学習する．

第4章 マルチタスクモデル

マルチタスク学習は自然言語処理の様々な分野に適用され、精度の向上が報告されている [28, 29, 30, 31, 32, 33, 34, 15]. 複数のタスクを同時に解くことで学習データが増え、各タスクに含まれるノイズに対して頑健でより良い表現を学習できるのがマルチタスク学習の利点の一つである. 本稿では、述語項構造解析と事態性名詞の項構造解析の2つのタスクに対して初めてマルチタスク学習を導入した. 我々は入力層, RNN層, 出力層においてシングルタスクモデルをマルチタスク学習モデルに拡張する3つの手法を提案する. 提案モデルを図4.1に示す. 最終的なモデルは3つ全ての手法を組み合わせたものである (図4.1e).

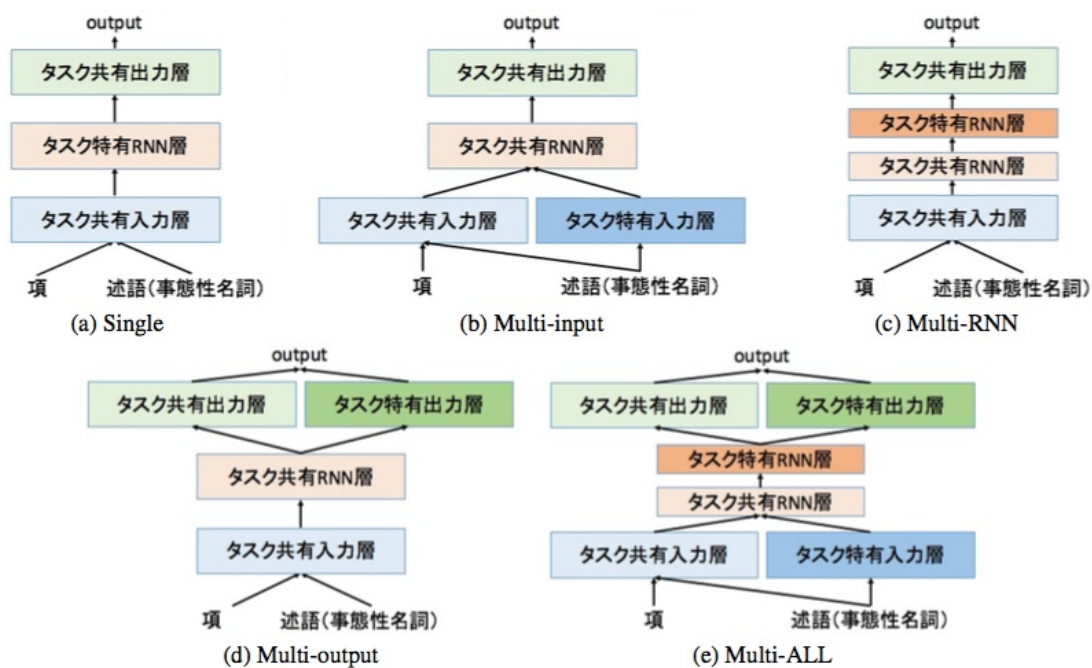


図 4.1: 提案モデル: (a) Single, (b) Multi-input, (c) Multi-RNN, (d) Multi-output, (e) Multi-ALL.

4.1 マルチ入力層

たとえ表層形が同じであっても，述語と事態性名詞では文脈が異なる．例えば図 1.1 の事態性名詞“報告”は述語と違い，同一文節内に項を持つ．また述語“短い”に対するガ格の項という役割も持っている．そこでこれらを考慮するために，我々はそのタスク独自の単語ベクトルを学習するタスク特有単語埋め込み行列を用意する．述語（事態性名詞） w_p はタスク特有単語埋め込み行列によってベクトル $\mathbf{x}_t^p, \mathbf{x}_t^n \in \mathbb{R}^{d_w}$ に変換される．

素性ベクトル $\bar{\mathbf{x}}_t$ は以下のように定義される．

$$\bar{\mathbf{x}}_t = \begin{cases} \mathbf{x}_t \oplus \mathbf{x}_t^p & (\text{PASA}) \\ \mathbf{x}_t \oplus \mathbf{x}_t^n & (\text{ENASA}) \end{cases} \quad (4.11)$$

PASA は述語項構造解析を，ENASA は事態性名詞の項構造解析を示す．シングルタスクモデルの入力層のみを拡張したモデルが図 4.1b である．

4.2 マルチ RNN 層

先行研究 [29, 31] では，容易なタスクから得られた特徴を困難なタスクに活用する階層的なマルチタスク学習のモデルが提案されている．これらの研究では，容易なタスクでは低層の RNN の表現を使い，困難なタスクでは高層の RNN の表現を利用することで予測精度が向上することが示された．そこで我々はタスク共有の RNN の上にタスク特有の RNN を階層的に重ねるネットワークを構築する．共有の RNN でタスク共有の知識表現を学習し，タスク特有の RNN でそれぞれのタスクに調整する．各タイムステップ t において， $l' \in [1, \dots, L']$ 層目の隠れ層ベクトル $\mathbf{m}_t^{l'} \in \mathbb{R}^{d_h}$ は以下のように計算される．

$$\mathbf{m}_t^{l'} = \begin{cases} g^{l'}(\mathbf{m}_t^{l'-1}, \mathbf{m}_{t-1}^{l'}) & (l' = \text{odd}) \\ g^{l'}(\mathbf{m}_t^{l'-1}, \mathbf{m}_{t+1}^{l'}) & (l' = \text{even}) \end{cases} \quad (4.21)$$

$$g^{l'}(\cdot) = \begin{cases} g_p^{l'}(\cdot) & (\text{PASA}) \\ g_n^{l'}(\cdot) & (\text{ENASA}) \end{cases} \quad (4.22)$$

$g^l(\cdot)$, $g_p^l(\cdot)$, $g_n^l(\cdot)$ は l 層目の GRU の処理を表す．また $\mathbf{m}_t^0 = \mathbf{h}_t^L$ である．シングルタスクモデルの RNN 層のみを拡張したモデルが図 4.1c である．

4.3 マルチ出力層

述語と事態性名詞では項の取り方が異なる．例えば，述語は同一文節内に項を持たないが，事態性名詞は複合名詞という形で同一文節内に項を持つ．出力層をタスク共有の層とタスク特有の層に分けるのは直感的で自然なことである．タスク特有の出力ベクトルは以下のように計算される．

$$\mathbf{o}_t^p = \mathbf{W}_o^p \mathbf{h}_t + \mathbf{b}_o^p \quad (4.31)$$

$$\mathbf{o}_t^n = \mathbf{W}_o^n \mathbf{h}_t + \mathbf{b}_o^n \quad (4.32)$$

$$\mathbf{g}_t = \sigma(\mathbf{W}_g \mathbf{h}_t + \mathbf{b}_g) \quad (4.33)$$

$\mathbf{W}_o^p, \mathbf{W}_o^n, \mathbf{W}_g \in \mathbb{R}^{4 \times d_h}$ はパラメータ行列， $\mathbf{b}_o^p, \mathbf{b}_o^n, \mathbf{b}_g \in \mathbb{R}^4$ はバイアス項である． \mathbf{h}_t は RNN の最終層の各隠れ層のベクトルである．タスク特有の出力ベクトル $\mathbf{o}_t^p, \mathbf{o}_t^n$ はゲート \mathbf{g}_t によってタスク共有の出力ベクトル \mathbf{o}_t と結合する．

$$\mathbf{c}_t = \begin{cases} \mathbf{g}_t \odot \mathbf{o}_t + (1 - \mathbf{g}_t) \odot \mathbf{o}_t^p & (\text{PASA}) \\ \mathbf{g}_t \odot \mathbf{o}_t + (1 - \mathbf{g}_t) \odot \mathbf{o}_t^n & (\text{ENASA}) \end{cases} \quad (4.34)$$

$$\bar{\mathbf{o}}_t = \text{softmax}(\mathbf{c}_t) \quad (4.35)$$

\odot は要素積を表す．出力ベクトル $\bar{\mathbf{o}}_t$ は，[ガ格, ヲ格, ニ格, ELSE] の確率を表すベクトルである．シングルタスクモデルの出力層のみを拡張したモデルが図 4.1d である．

第 5 章 実験

5.1 データセットと実験設定

実験データには NAIST テキストコーパス 1.5 [1] を用いた。Taira ら [3] と同様に、NAIST テキストコーパスの 1 月 1 日から 11 日の記事と 1 月から 8 月の社説を学習データに、NAIST テキストコーパスの 1 月 12 日から 13 日の記事と 9 月の社説を開発データに、NAIST テキストコーパスの 1 月 14 日から 17 日の記事と 10 月から 12 月の社説を評価データに使用した。単語境界と文節境界は NAIST テキストコーパスに付与されているデータを使用した。

開発・評価データについて、同じ必須格の項が文中に 2 つ以上ある場合の正解データの作り方は、述語と係り受けの関係にある項のみを正解とし、それ以外の項は ELSE ラベルにラベル付けする。係り受けの関係にある項がない場合は距離 $|w_p - w_t|$ が最短の項を正解とする。距離が等しい場合は述語より左側の項を正解とする。

NAIST テキストコーパス 1.5 では、“サ変接続名詞 + スル” のような述語があった場合に“スル”を述語単語としてアノテーションしている。しかし、述語と事態性名詞の表層形を一致させ、効果的に学習を行うために、我々はこの場合“サ変接続名詞”を述語単語として扱うように前処理を行った。例えば述語“報告する”と事態性名詞“報告”があった場合、前処理前の w_p はそれぞれ“する”、“報告”であるが、前処理後は“報告”に統一される。

5.2 ハイパーパラメータ

単語分散表現の初期値として朝日新聞単語ベクトル*の事前学習されたベクトルを使用した。タスク特有単語埋め込み行列は $[-0.25, 0.25]$ の一様分布に従ってサンプリングした値を初期値として設定した。学習データの中に 2 回以上出現する単語を単語ベクトルとし、それ以外の単語は未知語ベクトルとして扱う。最適化手法には AdaDelta ($\epsilon = 10^{-6}$, $\rho = 0.95$) を適用した。エポック数は 20 に設定し、開

*http://www.asahi.com/shimbun/medialab/word_embedding

表 5.1: ハイパーパラメータ

単語の埋め込みベクトルの次元数 d_w	300
位置埋め込みベクトルの次元数 d_p	16
係り受け関係埋め込みベクトルの次元数 d_d	16
GRU の隠れ層の次元数 d_h	300
GRU の層数 L	4
タスク特有単語の埋め込みベクトルの次元数 d'_w	16
タスク特有 GRU の隠れ層の次元数 d'_h	300
タスク特有 GRU の層数 L'	2
dropout rate	0.4
batch size	8
gradient clipping	4

発データの F 値が最も高かったエポックのモデルで評価データを評価する。ハイパーパラメータを表 5.1 に示す。

5.3 実験結果

NAIST テキストコーパス 1.5 の評価データで各モデルの評価を行った。文内述語項構造解析と文内事態性名詞の項構造解析の実験結果を表 5.2, 5.3 に示す。2 列目に全体の F 値と標準偏差を示し、3 列目以降は項カテゴリ別の F 値を記載した。

5.3.1 述語項構造解析

表 5.2 の 1 行目に先行研究のモデルの結果をまとめた。Ouchi+ 17 は Ouchi ら [10] の Multi-Seq モデルである。M&I 17 は Matsubayashi と Inui [9] のモデルである。M&I 18 は Matsubayashi と Inui [12] の MP-POOL-SELFATT モデルである。

2 行目に提案手法のモデルの結果をまとめた。これらのモデルでは述語と項の係り受け関係を示す素性は用いていない。Single と Multi-input, Multi-RNN, Multi-output, Multi-ALL をそれぞれ比較すると、どのモデルでも全体の F 値が向

表 5.2: 評価データにおける述語項構造解析の F1 スコア

Method			Dep				Zero			
	ALL	SD	ALL	ガ格	ヲ格	ニ格	ALL	ガ格	ヲ格	ニ格
Ouchi+ 17	81.42		88.17	88.75	93.68	64.38	47.12	50.65	32.35	7.52
M&I 17	83.50	±0.17	89.89	91.19	95.18	61.90	51.79	54.69	41.8	17
M&I 18	83.94	±0.12	90.26	90.88	94.99	67.57	55.55	57.99	48.9	23
Single	83.62	±0.17	90.09	90.45	94.84	69.77	51.87	54.73	43.48	11.40
Multi-input	83.88	±0.11	90.27	90.65	95.12	69.86	53.01	55.82	44.68	10.77
Multi-RNN	83.91	±0.23	90.17	90.58	95.07	67.94	53.31	55.85	45.71	9.97
Multi-output	83.77	±0.20	90.13	90.68	94.89	68.16	53.93	56.73	43.79	9.45
Multi-ALL	83.82	±0.10	90.15	90.68	95.06	67.56	53.50	56.37	45.36	8.70
Multi-RNN+DEP	84.55	±0.11	90.69	91.28	95.25	70.07	51.56	54.29	42.67	1.85
Multi-output+DEP	84.73	±0.11	90.82	91.46	95.29	70.69	52.29	55.14	42.15	1.81
Multi-ALL+DEP	84.75	±0.16	90.88	91.40	95.37	71.02	52.35	55.10	42.54	2.32
M&I 17 (ens. of 5)	84.07		90.24	91.59	95.29	62.61	53.66	56.47	44.7	16
M&I 18 (ens. of 10)	85.34		91.26	91.84	95.57	70.8	58.07	60.21	52.5	26
Multi-RNN+DEP (ens. of 5)	85.85		91.61	92.11	95.87	72.63	53.41	55.96	46.10	0
Multi-output+DEP (ens. of 5)	85.83		91.52	92.12	95.69	72.72	54.35	57.02	45.95	0
Multi-ALL+DEP (ens. of 5)	86.01		91.63	92.15	95.80	72.95	54.99	57.84	45.20	0

表 5.3: 評価データにおける事態性名詞の項構造解析の F1 スコア

Method			Dep				Zero				Bunsetsu					
	ALL	SD	ALL	ガ格	ヲ格	ニ格	ALL	ガ格	ヲ格	ニ格	ALL	ガ格	ヲ格	ニ格		
Taira+ 08 on NTC 1.4				68.01	62.46	56.05			36.19	20.46	6.62			78.93	77.96	58.13
Single	66.21	±0.15	74.64	76.06	74.54	51.28	46.05	49.67	33.36	13.63	78.24	76.67	81.75	48.55		
Multi-input	67.89	±0.42	75.62	76.63	75.78	57.17	49.07	52.81	36.95	19.39	79.35	77.31	83.31	51.03		
Multi-RNN	67.96	±0.44	75.86	76.90	76.33	54.46	48.67	52.18	38.47	18.89	79.08	77.24	82.89	50.93		
Multi-output	67.96	±0.17	76.25	77.18	76.90	54.97	48.74	52.48	36.09	19.64	79.02	77.00	83.04	50.60		
Multi-ALL	68.00	±0.41	75.90	77.16	76.05	53.00	49.66	53.37	37.64	14.46	79.05	77.32	82.61	51.83		
Multi-ALL+DEP	67.68	±0.39	75.95	77.18	76.11	55.26	47.57	51.21	35.14	15.65	79.06	77.44	82.66	51.10		
Multi-ALL (ens. of 5)	71.14		78.63	79.66	78.83	58.29	52.49	56.41	39.02	16.42	81.90	80.25	85.21	56.29		
Multi-ALL+DEP (ens. of 5)	69.90		77.86	78.89	78.16	58.46	49.36	53.10	36.36	17.23	81.16	79.74	84.57	52.99		

上することを示した。その中でも Multi-RNN が Single と比較して全体の F 値で 0.29 ポイントの精度向上を示した。Ouchi ら [10] や Matsubayashi と Inui [12] は複数の述語項関係を考慮することで、全体の F 値でそれぞれ 0.27, 0.55 ポイントの精度向上を示している。このことから事態性名詞の項構造解析のマルチタスク学習がこれらの研究と同程度の効果を持つことを示すことができた。

3 行目に係り受け素性を含めた全ての素性を用いた提案手法のモデルの結果をまとめた。Multi-ALL+DEP が先行研究を含めた全てのモデルの中で最も高い解析精

度を示した。特に Dep の事例において係り受け素性の効果が高いことが分かった。一方で Zero の事例においては解析精度が低く、事例数の多い Dep を優先した最適化を行っていることが分かった。

4 行目にアンサンブルしたモデルの結果をまとめた。最終的に、我々の提案モデルは先行研究を 0.67 ポイント上回る解析精度を示した。

またこのモデルは文内の述語の項推定を独立に行う単純なモデルであるが、Ouchi ら [10] や Matsubayashi と Inui [12] のモデルより高い解析精度を示している。近年では、複数の述語を同時に考慮し文内単位で最適化を行う手法が盛んに研究されてきたが、end-to-end モデルにおける統語情報の活用方法も今後研究されるべき課題の一つである。

5.3.2 事態性名詞の項構造解析

表 5.3 の 1 列目に先行研究のモデルの結果をまとめた。Taira+ 08 は Taira ら [3] のモデルである。このスコアは NAIST テキストコーパス 1.4 での実験結果であるため、提案手法との厳密な比較はできない。Single と Multi-input, Multi-RNN, Multi-output, Multi-ALL をそれぞれ比較すると、どのモデルでも全体の F 値が向上することを示した。一方述語項構造解析で最高精度を示した Multi-ALL+DEP は Multi-ALL を下回る精度を示した。このことから、項述語間の係り受け関係は事態性名詞の項構造解析において有効な素性でないことが分かった。

第6章 分析

表 6.1: 評価データにおける述語項構造解析の誤り事例

(a)	社会党の山花貞夫・新民主連合会長は十七日午前、新たに結成する新会派に参加する同党所属国会議員二十四人の「日本社会党・護憲民主連合」からの離島届を森井忠良国対委員長に提出した。	
	Multi-RNN	Single
	ガ格	会長
	ヲ格	会派
(b)	行政、銀行、農協系統の経営トップが責任回避するのはなぜか。	
	Multi-RNN	Single
	ガ格	トップ
	ヲ格	責任
(c)	私は、紙幣を改めて見ると、なるほど千円札の左下に直径五ミリほどの丸が一つ押されてある。	
	Single	Multi-RNN
	ガ格	-
	ヲ格	丸
(d)	ローズ博士は、アポEが発病に果たす役割の解明に取り組んでおり、「発病を予防する薬の開発をめざしている」という。	
	Multi-ALL+DEP	Multi-ALL
	ガ格	E
	ヲ格	役割
	ニ格	発病

提案モデルの効果を分析するために、評価データにおいて各モデルの述語項構造解析の出力結果を比較したものを表 6.1 に示す。(a) と (b) の例は Multi-RNN モデルでは正しく予測できたが、Single モデルでは誤った予測を出力した事例である。(a) では述語“結成する”に対して、“会長”がガ格、“会派”がヲ格である。しかしながら、Single モデルはどちらの格もこの文中に存在しないと間違って出力している。各タスクの学習データを比較すると、文内事態性名詞の項構造解析の事例

表 6.2: 評価データにおける連体修飾に対する述語項構造解析の F1 スコア

	ALL	ガ格	ヲ格	ニ格
Multi-ALL	80.31	83.37	72.16	19.48
Multi-ALL+DEP	81.83	84.67	74.41	28.31

数は文内述語項構造解析の事例数の約 3 分の 1 ほどであるにも関わらず、事態性名詞“結成”を対象にした事例数は述語“結成する”を対象にした事例数の約 2 倍であった。このことから、Multi-RNN モデルはマルチタスク学習によって効果的に事態性名詞の情報を活用できていることが分かる。(b) では述語“責任回避する”に対して、“トップ”がガ格、同一文節内の“責任”がヲ格である。これは述語と同一文節内に項が存在する興味深い事例であり、この項は評価対象から除外されているが、事態性名詞では複合名詞としてよく見られる。そのため、Single モデルではヲ格の項は存在しないと誤って予測しているが、Multi-RNN では事態性名詞の知識を利用することで正しく予測できている。

反対に、(c) の例は Single モデルでは正しく予測できたが、Multi-RNN モデルでは誤った予測結果を返した事例である。述語“押す”に対して、ヲ格に関してはどちらのモデルも正しく出力できているが、ニ格に関しては Multi-RNN モデルが「項なし」と誤って出力している。この事例のように、項の直後に与格を示す助詞の“に”があるにも関わらず、Multi-RNN が誤ったニ格を出力する事例が多く見られた。このことから、ニ格の解析については、事態性名詞の情報が述語項構造解析に悪影響を与えていることが分かる。

次に、(d) の例は Multi-ALL+DEP モデルでは正しく予測できたが、Multi-ALL モデルでは誤った予測をした事例である。述語“果たす”に対して、Multi-ALL+DEP モデルは係り受け関係を持つ項“役割”をヲ格と予測できているが、Multi-ALL モデルは“解明”をヲ格の項と間違えて予測している。このような連体修飾が含まれる事例において、統語情報を用いない Multi-ALL モデルの解析ミスが多く確認された。表 6.2 は連体修飾に対する 2 つのモデルの述語項構造解析結果である。係り受け関係を素性として用いることで、全ての必須格において解析精度が大きく向上していることが分かる。このことから、特に連体修飾を持つ述語項構造解析において係り受けタイプ素性が有効であることが示された。

第7章 おわりに

本論文では，述語項構造解析と事態性名詞の項構造解析をマルチタスク学習するモデルを提案した．先行研究は述語のみを対象としたものがほとんどであり，事態性名詞を除いて解析を行う，もしくは単一のモデルで述語と事態性名詞の解析を行っていた．しかし，事態性名詞は述語と同様に項を持つため，正しく文章の意味構造を解析するうえで，述語のみを対象にした項構造解析は不十分である．本研究では，マルチタスク学習によってタスク間で共通の特徴とタスク特有の特徴を区別し，効果的に述語と事態性名詞の情報を活用できると考え，実験を行った．その結果 NAIST テキストコーパスにおいて双方のタスクで解析精度が向上することを示した．そして，係り受け関係を素性として加えることによって，文内述語項構造解析において最高精度を達成した．また，本研究は深層学習を用いて事態性名詞の項構造解析を行った初めての研究である．今後の展望としては，複数の述語と事態性名詞を考慮したモデルへの改良などが挙げられる．

発表リスト

1. 大森光, 小町守. 複数の項候補を考慮したニューラル文内述語項構造解析.
言語処理学会第 24 回年次大会, pp.252-255. March 13, 2018.

謝辞

本論文の執筆に際して，研究室に配属されてから3年間ご指導ご鞭撻を頂きました小町守准教授に深く感謝いたします。また日々の研究生活の中で，研究室の同期，先輩，後輩には，多くのアドバイスやご協力を頂きました。また副査を引き受けてくださった山口教授，高間教授に心より感謝いたします。皆様へ心から感謝の気持ちと御礼を申し上げたく，謝辞に返させていただきます。

参考文献

- [1] R. Iida, M. Komachi, K. Inui, and Y. Matsumoto, “Annotating a Japanese text corpus with predicate-argument and coreference relations,” Proceedings of the Linguistic Annotation Workshop, pp.132–139, 2007.
- [2] M. Komachi, R. Iida, K. Inui, and Y. Matsumoto, “Learning based argument structure analysis of event-nouns in Japanese,” Proceedings of PACLING, pp.120–128, 2007.
- [3] H. Taira, S. Fujita, and M. Nagata, “A Japanese predicate argument structure analysis using decision lists,” Proceedings of EMNLP, pp.523–532, 2008.
- [4] K. Imamura, K. Saito, and T. Izumi, “Discriminative approach to predicate-argument structure analysis with zero-anaphora resolution,” Proceedings of ACL-IJCNLP, pp.85–88, 2009.
- [5] Y. Hayashibe, M. Komachi, and Y. Matsumoto, “Japanese predicate argument structure analysis exploiting argument position and type,” Proceedings of IJCNLP, pp.201–209, 2011.
- [6] K. Yoshikawa, M. Asahara, and Y. Matsumoto, “Jointly extracting Japanese predicate-argument relation with Markov logic,” Proceedings of IJCNLP, pp.1125–1133, 2011.
- [7] H. Ouchi, H. Shindo, K. Duh, and Y. Matsumoto, “Joint case argument identification for Japanese predicate argument structure analysis,” Proceedings of ACL-IJCNLP, pp.961–970, 2015.
- [8] T. Shibata, D. Kawahara, and S. Kurohashi, “Neural network-based model for Japanese predicate argument structure analysis,” Proceedings of ACL, pp.1235–1244, 2016.
- [9] Y. Matsubayashi and K. Inui, “Revisiting the design issues of local models for Japanese predicate-argument structure analysis,” Proceedings of IJCNLP, pp.128–133, 2017.
- [10] H. Ouchi, H. Shindo, and Y. Matsumoto, “Neural modeling of multi-predicate interactions for Japanese predicate argument structure analysis,” Proceedings of ACL, pp.1591–1600, 2017.
- [11] J. Zhou and W. Xu, “End-to-end learning of semantic role labeling using recurrent neural networks,” Proceedings of ACL-IJCNLP, pp.1127–1137, 2015.
- [12] Y. Matsubayashi and K. Inui, “Distance-free modeling of multi-predicate interactions in end-to-end Japanese predicate argument structure analysis,” Proceedings of COLING, pp.94–106, 2018.
- [13] L. He, K. Lee, M. Lewis, and L. Zettlemoyer, “Deep semantic role labeling: What

- works and what’s next,” Proceedings of ACL, pp.473–483, 2017.
- [14] Z. Tan, M. Wang, J. Xie, Y. Chen, and X. Shi, “Deep semantic role labeling with self-attention,” Proceedings of AAAI, pp.4929–4936, 2018.
 - [15] E. Strubell, P. Verga, D. Andor, D. Weiss, and A. McCallum, “Linguistically-informed self-attention for semantic role labeling,” Proceedings of EMNLP, pp.5027–5038, 2018.
 - [16] H. Ouchi, H. Shindo, and Y. Matsumoto, “A span selection model for semantic role labeling,” Proceedings of EMNLP, pp.1630–1642, 2018.
 - [17] M. Gerber and J. Chai, “Beyond NomBank: A study of implicit arguments for nominal predicates,” Proceedings of ACL, pp.1583–1592, 2010.
 - [18] Z. Wang, T. Jiang, B. Chang, and Z. Sui, “Chinese semantic role labeling with bidirectional recurrent neural networks,” Proceedings of EMNLP, pp.1626–1631, 2015.
 - [19] M. Roth and M. Lapata, “Neural semantic role labeling with dependency path embeddings,” Proceedings of ACL, pp.1192–1202, 2016.
 - [20] L. Sha, S. Li, B. Chang, Z. Sui, and T. Jiang, “Capturing argument relationship for Chinese semantic role labeling,” Proceedings of EMNLP, pp.2011–2016, 2016.
 - [21] D. Marcheggiani, A. Frolov, and I. Titov, “A simple and accurate syntax-agnostic neural model for dependency-based semantic role labeling,” Proceedings of CoNLL, pp.411–420, 2017.
 - [22] F. Qian, L. Sha, B. Chang, L. Liu, and M. Zhang, “Syntax aware LSTM model for semantic role labeling,” Proceedings of the 2nd Workshop on Structured Prediction for Natural Language Processing, pp.27–32, 2017.
 - [23] J. Li, G. Zhou, H. Zhao, Q. Zhu, and P. Qian, “Improving nominal SRL in Chinese language with verbal SRL information and automatic predicate recognition,” Proceedings of EMNLP, pp.1280–1288, 2009.
 - [24] R. Iida, K. Inui, and Y. Matsumoto, “Exploiting syntactic patterns as clues in zero-anaphora resolution,” Proceedings of COLING/ACL, pp.625–632, 2006.
 - [25] R. Sasano and S. Kurohashi, “A discriminative approach to Japanese zero anaphora resolution with large-scale lexicalized case frames,” Proceedings of IJCNLP, pp.758–766, 2011.
 - [26] K. Cho, B. van Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, “Learning phrase representations using RNN encoder–decoder for statistical machine translation,” Proceedings of EMNLP, pp.1724–1734, 2014.
 - [27] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” Proceedings of CVPR, pp.770–778, 2016.

- [28] R. Collobert, J. Weston, L. Bottou, M. Karlen, K. Kavukcuoglu, and P. Kuksa, “Natural language processing (almost) from scratch,” *J. Mach. Learn. Res.*, vol.12, pp.2493–2537, 2011.
- [29] A. Søgaard and Y. Goldberg, “Deep multi-task learning with low level tasks supervised at lower layers,” *Proceedings of ACL*, pp.231–235, 2016.
- [30] M. Luong, Q.V. Le, I. Sutskever, O. Vinyals, and L. Kaiser, “Multi-task sequence to sequence learning,” *Proceedings of ICLR*, pp.●●–●●, 2016.
- [31] K. Hashimoto, c. xiong, Y. Tsuruoka, and R. Socher, “A joint many-task model: Growing a neural network for multiple NLP tasks,” *Proceedings of EMNLP*, pp.1923–1933, 2017.
- [32] P. Liu, X. Qiu, and X. Huang, “Adversarial multi-task learning for text classification,” *Proceedings of ACL*, pp.1–10, 2017.
- [33] V. Stoyanov, H. Ji, Y. Lin, and S. Yang, “A multi-lingual multi-task architecture for low-resource sequence labeling,” *Proceedings of ACL*, pp.799–809, 2018.
- [34] A. Marasovic and A. Frank, “SRL4ORL: Improving opinion role labeling using multi-task learning with semantic role labeling,” *Proceedings of NAACL*, pp.583–594, 2018.